DOCUMENT RESUME

ED 326 072 FL 018 999

AUTHOR Virkkunen, Anu
TITLE The Life of a Test or Can a Discrete-Item Test Measure a Construct?
PUB DATE Apr 90
NOTE 11p.; Paper presented at the Regional Language Centre Seminar (Singapore, April 9-12, 1990).
PUB TYPE Speeches/Conference Papers (150) -- Tests/Evaluation Instruments (160)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Affixes; Comparative Analysis; *Construct Validity; *English (Second Language); Foreign Countries; Higher Education; *Language Tests; *Reading Comprehension; Second Language Instruction; Test Construction; *Test Format
IDENTIFIERS *University of Helsinki (Finland)

ABSTRACT

A study investigated whether or not a discrete-item test of English affixes could be used to measure second language reading comprehension. Subjects were 1,254 mostly first-year university students in four academic departments studying to pass the first half of a foreign language requirement, an English reading comprehension test. Results from two affix tests were compared. The first was the test in use at the University of Helsinki (Finland) for a period of years. The second test was designed for the experiment, based on material covered in the course, developed by a non-native English-speaker, edited by a native speaker, and revised after early trials. The original test was administered in September 1986, the second in December 1986, both with other reading comprehension measures. Analysis of results indicates that the two tests were internally consistent, and that the affix tests correlated more closely with cloze than with multiple-choice or open-ended questions. However, the results also showed that the two consecutive but different affix tests were not equally valid. Because of experiment limitations, further research is recommended. The tests are appended. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

presented at the Regional Language Centre Seminar
April 9-12, 1990, Singapore

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
X This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
A. Virkkunen
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ED326072

ANU VIRKKUNEN
UNIVERSITY OF HELSINKI

THE LIFE OF A TEST or CAN A DISCRETE-ITEM TEST MEASURE A CONSTRUCT?

Introduction

It all started with test anxiety.

Every student at the University of Helsinki has to pass foreign language requirements. As the most popular foreign language is English in the Finnish school system (more than 90% of all Finnish students start studying English as their first foreign language in the third grade), English is also the most popular language at the universities.

The foreign language requirements every student has to pass in at least one language, have two parts. The first part is a reading comprehension test based on subject-specific materials in the student's own area of studies. The second part is an oral-aural test in the same language. My area of testing as a non-native English teacher is English reading comprehension (ERC).

The English reading comprehension test consists of several components. The common set-up is a combination of one semantic cloze test and two tests with either open-ended or multiple choice questions. Now, lately, our teachers have been experimenting with a precis as part of the test. Anyway, the student taking the test knows that all the texts or materials the test is based on will somehow deal with the area of his studies but that he will always encounter new material in the test.

The students who fail the test when they take it for the first time have to attend a reading comprehension course before they can take the test again. One component of the course is the teaching of affixes. Teaching morphological analysis of words to speakers of a non-Indo-European language like Finnish has proved very beneficial for two reasons. Firstly, morphological derivation is a word formation technique that is also widely used in Finnish. Thus, as a technique, the Finns are already familiar with it and understand its functions. Secondly, even such frequent prefixes as non- and super- are non-existent as loan words or cognates in Finnish. They have Finnish equivalents: super- is yli- in Finnish, but non- is expressed usually with a suffix. However they are expressed, with suffixes or prefixes, once the Finnish student knows of the meaning of the most frequent English prefixes and suffixes, he can cope with morphological analysis quite well in English.

The students who had failed the test, attended the course and were facing a second ERC test, brought up the question of test anxiety. Their suggestion was to include a component in the test that they could study for. As the ERC courses consist of reading subject-specific texts, learning different reading strategies, like morphological decomposition, and getting familiar with the most common structural features of academic writing,

FL018499

affixes seemed to best lend themselves for testing. A short test component you could study for and start the test with but only get one fifth of the total final points at the maximum, was regarded as a good solution. The students were happy, and the affix test results seemed to correlate quite well with the reasults of the other test components.

However, with the time, ERC teachers started questioning whether the recognition of affixes can be regarded as a measure of ERC. This paper will first deal shortly with the development of the test, and then explain the study that was conducted to find out if a discrete-item affix test can actually be used to measure reading comprehension (the construct). The basic structure of this paper follows the four phases in the devising of a test suggested by Carroll and Hall (1985). With reliability and validity, also efficiency of the test will be discussed (Weir 1988).

PHASE I PLANNING

Theory

In cognitive psychology, affixes are discussed as part of the mental lexikon (Just and Carpenter 1987). The question concerning affixes is whether prefixed and suffixed words are arranged with their affixes in the mental lexikon. If it were true that e.g. prefixed words are grouped together according to their prefixes, then the prefixes would be important organizators of the mental lexikon. There are two major schools of thought. There are those who support the Taft and Forster model (1975), which is also called the prefix-stripping model. Taft and Forster, in their study, showed that prefixes are not stored with the words they combine with but are stored somewhere in the vicinity. Since then, Taft has conducted several studies to confirm the prefix-stripping model. Taft's results with suffixes were more ambiguous. He does suggest, though, that suffixes are also stripped and stored near the words they combine with.

The second school of thought was started with a study by Rubin, Becker and Freeman (1979). The study was based on the Taft and Forster model but their conclusions were different. They regard prefix-stripping, or morphological decomposition, as it is also called, a plausible strategy but rather exceptional than common.

Another approach to affixation is the Russian theory of a potential vocabulary (Berman, Buchbinder and Bezdeneznych 1968). Side by side with the active and passive vocabularies, there is a potential vocabulary that consists of words one has never seen but will comprehend at first sight. An example of this is the word "unlike". A native speaker of English who has never seen the word before can understand it if he knows the prefix "un" from e.g. "unhappy" and the word "like". Berman, Buchbinder and Bezdeneznych advocate teaching affixes as part of foreign language curricula. They even refer to fabulous increases in vocabulary after just a couple of affixes have been learnt. This idea was followed by e.g.

Denninghaus, who also showed in his studies (1976) how students could enlarge their vocabularies.

There was very little research or any other theoretical material on testing affixes. We did find out about one test on affixes, "Morpheme Knowledge Test", which had beem devised by James F. Shepherd in 1973. However, there has been no mention of how, where and when this test has been used. Nor was their any mention of the results the test had yielded.

Testees

A typical testee was a first year University of Helsinki student studying to pass the first half of a foreign language requirement, the reading comprehension test in English.

Needs analysis

The affix test would clearly be an answer to the voiced need to have a part of the test you could study for. The test would also base on material dealt with during the ERC course.

Subject areas and tasks

Though our students in Medicine and Biology have their own affix lists and books, it was decided in the first phase that the so-called "general" affixes be focused on. It was also decided that the type of item is a one- or two-clause sentence context and a target word with the suffix or the prefix deleted. In the first version, four alternatives were provided for every item. The a priori validation was guaranteed by checking first that all the affixes included in the test, were from the Finnish textbook on affixes the students had studied during the course.

PHASE II Developing the test

The first drafts were devised by a non-native English teacher. Then the test items were checked by a native speaker teacher of English. That was done because we wanted to make sure that there was only one correct answer, and because the contexts had to appear natural.

After the first experiments with the test, certain practical improvements were made. The number of test items was decided to be twenty: either ten items with deleted prefixes and another ten items with deleted suffixes, or fifteen items with deleted prefixes and five with deleted suffixes. Instead of the four alternatives to one item, the items were grouped into groups of five. Every group was preceded by ten plausible prefixes or suffixes. Every affix was supposed to be used only once, which fact was not mentioned to the students. Also the native-speaker teacher was to make sure that only one of the given alternatives was to be correct.

As the affix test was only used at the University of Helsinki, any kind of standardization was out of question. A new affix test on the same lines was devised for every ERC test.

Another practical improvement was the development of the answer sheet, which made the correcting of papers much faster. This point of efficiency was very important because seven teachers had to correct about 3 000 ERC tests in about a week (see Appendices 1 and 2).

PHASE III Administration of the test

The improved affix test was made part of the ERC test in 1982. The description of the ERC test in the Language Centre Program, a booklet published and distributed to the students, mentioned the affix test and also gave the name of the Finnish textbook on affixes. The students in the ERC courses were also notified. In the beginning of every ERC test, students were given general guidelines as to how much time they should spend on the different parts of the test. The time they would need for completing the affix test was calculated to be twenty minutes. As the students got all the papers at the same time, whether they spent twenty minutes on the affix test or more was not checked.

PHASE IV Follow-up

In use, the affix test proved quite effective--with the conversion of correct answers to score points, one teacher corrected almost a paper a minute. The test had face validity, the students were quite happy with it. But did the test really measure ERC? The test had been used for four years when it was decided that a study be done to show what the test actually measured. It was also decided that no more affix tests be used before there are research results to rely on. I undertook the study and it became my doctoral dissertation at the University of Minnesota.

ABOUT THE STUDY

As the two affix tests administered as part of the ERC tests in September and December 1986 were the last affix tests used, it was decided that the data for this study come from those tests. Four faculties were chosen: 1) the Faculty of Agriculture and Forestry, 2) the Faculty of Law, 3) the Faculty of Mathematics and Sciences, and 4) the Faculty of Theology. All together, the study had 1254 subjects from those four faculties. Of the subjects, their total scores, test component scores and their affix test results were known as were also known their sex and age.

The following research questions formed the basis of the study:
1) Does the given sentence context influence mastery of affixation, and if so, how? Do context clues facilitate students' choice of affixes?
2) Is the Finnish test valid and reliable? Do all the test items measure the same construct? How does the Finnish affix test relate to the other reading comprehension measures (semantic cloze tests, and tests with either

multiple-choice or open-ended questions), and does it yield reliable results?
3) Are two consecutive, but different affix tests equally valid?
4) Does formal study of affixes predict better results in the affix test? Do students improve their scores on an affix test after formal study?

To find out about the context influence, item analysis was carried out. Content validity was checked, and the calculation of context clues was done. Also a linguistic analysis of the test items was carried out. The surprising result was that cuing was not a facilitating factor. Word frequency was not taken into account in this study, as it had been found out in a previous study on the reproduction of negative prefixes that word frequency had nothing to do with the results (Dotzenroth and Virkkunen, forthcoming). But perhaps it should have been included. Item difficulty ranged between .27 and .98. Item discriminability varied between .87 and .51, the lowest indices being for the easiest test items (p =.98). An interesting implication for further study came up in the analysis of the answer frequencies. In some test items, e.g. the most difficult test item, number 19 in the September test, the students had understood the cuing correctly but chosen the wrong adjectival endings for the term 'scholarly', as the suffix -ly is the most common adverbial ending. In other words, with tricky items like that, cuing can also function as a distractor. In further study, attention must be paid to the choice of distractors.

To establish the internal consistency of the test items in these two affix tests, Cronbach's alpha was calculated. It was .78 but high enough to prove that the test items measure the same thing. To establish construct validity, Pearson's product-moment correlations were calculated between the different test types and age. The affix tests correlated more highly with the cloze tests than with the multiple-choice or open-ended questions. The correlations varied between .77 and .52 (affix test and cloze), between .66 and .36 (affix test and multiple-choice) and between .53 and .32 (affix test and open-ended questions). To support those correlations, correlations were also calculated between the subjects' ages and affix test results, and they correlated either nonsignificantly with higher indices, e.g. .20 with p=.06, or with significant but low indices, e.g. .14 with p=.000. Here it was taken for granted that all the other measures were true measures of ERC. The problem with the correlations between different test components was, however, that both the September and December tests included several tests of the same kind, even in the same faculty, as students were grouped according to their major subject or area of study. Therefore, for example the cloze tests had to be taken as a genre rather than as separate tests in a pretest-posttest situation. Therefore, to be really able to establish construct validity, more research has to be done.

On the other hand, based on these data, the results show that two consecutive, but different affix tests are not equally valid. The results were arrived at calculating correlation coefficients for all Law students' scores in the September and December tests. The Pearson product-moment

correlation coefficient between the two affix tests was only .35. Next, the results from the Faculty of Law were used. The September and December correlation coefficients between the affix tests and the cloze tests; and the null hypothesis was that these coefficients would be equal. The null hypothesis had to be rejected as the value of the test statistic, based on Fisher's transformation (Ranta, Rita and Kouki 1989), was z=3.48, which is highly significant (p=.03%).

The fourth research question dealt with the influence of formal study on affix test results. Subjects did improve their test results quite significantly. Fisher's t-test yielded t= -4.53 for 139 subjects, with p=.000.

DISCUSSION

The study showed quite clearly the considerable correlation between an affix test and other measures of reading. However, because the data came from two authentic tests and the numbers of subjects were so great, this research could not be turned into a neat, small study with clear-cut and easily interpretable results. Rather, it crystallized many questions for further study. It is obvious that more research has to be done before we can decide whether an affix test can be used as a measure of ERC, but so far, there does not seem to be reason not to use it. The higher correlations with the cloze suggest that it also measures the reading process. As a strategy, morphological decomposition seems to be very beneficial for a student of English with a non-Indo-European language background. How beneficial, more research into the concept of potential vocabulary might show. It would also be very interesting to do research with English teachers in Asia and study the effects of different writing styles. Would teaching affixation be equally motivating for a Japanese or a Korean or a Malay student?

We will continue studying affixation and reading comprehension. The next area will be Medical English, where we wish to have a regular pretest-posttest set-up.

REFERENCES

Berman, I.M., V.A. Buchbinder and M.L. Bezdeneznych. "Formirovanie potentialnogo slovarnogo zapasa pri obucenii russkomu jazyku kak inostrannomu (Formation of a potential reserve vocabulary in learning Russian as a foreign language)." Russkij jazyk za rubezom, 4, 1968: 57-60.

Carroll, Brendan J., and Patrick J. Hall. Make Your Own Language Tests. A practical guide to writing language performance tests. Language Teaching Methodology Series. Oxford: Pergamon Press Ltd., 1985.

Denninghaus, Friedhelm. "Der kontrollierte Erwerb eines potentiellen Wortschatzes im Fremdsprachenunterricht." Praxis des neu-sprachlichen Unterrichts, 1, 1976: 3-14.

Just, Marcel Adam, and Patricia A. Carpenter. The Psychology of Reading and Language Comprehension. Boston: Allyn and Bacon, 1987.

Ranta, Esa, Hannu Rita and Jari Kouki. Biometria tilastotiedettä ekologeille. 2nd revised ed. Helsinki: Yliopistopaino, 1989.

Rubin, Gary S., and Curtis A. Becker. "Morphological Structure and Its Effect on Visual Word Recognition." Journal of Verbal Learning and Verbal Behavior, 18, 1979: 757-67.

Shepherd, James F. Morpheme Knowledge Test. The Research Instruments Project (TRIP), 1973.

Taft, Marcus, and Kenneth I. Forster. "Lexical Storage and Retrieval of Prefixed Words." Journal of Verbal Learning and Verbal Behavior, vol. 14, no.6, 1975: 638-47.

Weir, Cyril J. Communicative Language Testing with Special Reference to English as a Foreign Language. Exeter Linguistic Studies Vol. 11. Exeter: A. Wheaton & Co Ltd., 1988.

CHOOSE THE SUITABLE AFFIX AND MARK IT DOWN ON THE SEPARATE ANSWER SHEET

a) de-	b) dis-	c) trans-	d) mal-	e) non-	f) bi-	g) en-		
h) uni-	i) il-	j) out-						

1. In Finland there are many _____lingual people. They speak Finnish and Swedish equally well.
2. He _____approved of your behaviour. I also think it wasn't nice.
3. Literature other than novels, poems, plays and stories is called _____-fiction.
4. A _____continental railway crosses a continent.
5. It's _____legal to park your car here.

a) sub-	b) mis-	c) pro-	d) co-	e) ir-	f) mono-	
g) counter-	h) semi-	i) multi-	j) un-			

6. Your button has come _____done. Can I button it up for you?
7. I missed the train; I had _____calculated the time it would take to reach the station.
8. A house which is joined to another on one side only is called a _____-detached house.
9. In some countries the tobacco industry is a government _____poly. Only the government has the right to make cigarettes.
10. The workers, in _____-operation with the management, have increased output by 10 per cent.

a) -hood	b) -dom	c) -ship	d) -y	e) -ous	f) -like	
g) -some	h) -age	i) -ish	j) -ful			

11. He was so cold that even his lips were blu(e)_____.
12. Thank you for having been so help_____.
13. My grandfather often told us stories about his boy_____.
14. It took some good salesman_____ to get his plan accepted by the boss.
15. She was full of anger and gave him a murder____ look.

a) -worthy	b) -less	c) -hood	d) -ify	e) -ize	f) -ly
g) -ment	h) -ness	i) -ship	j) -y		

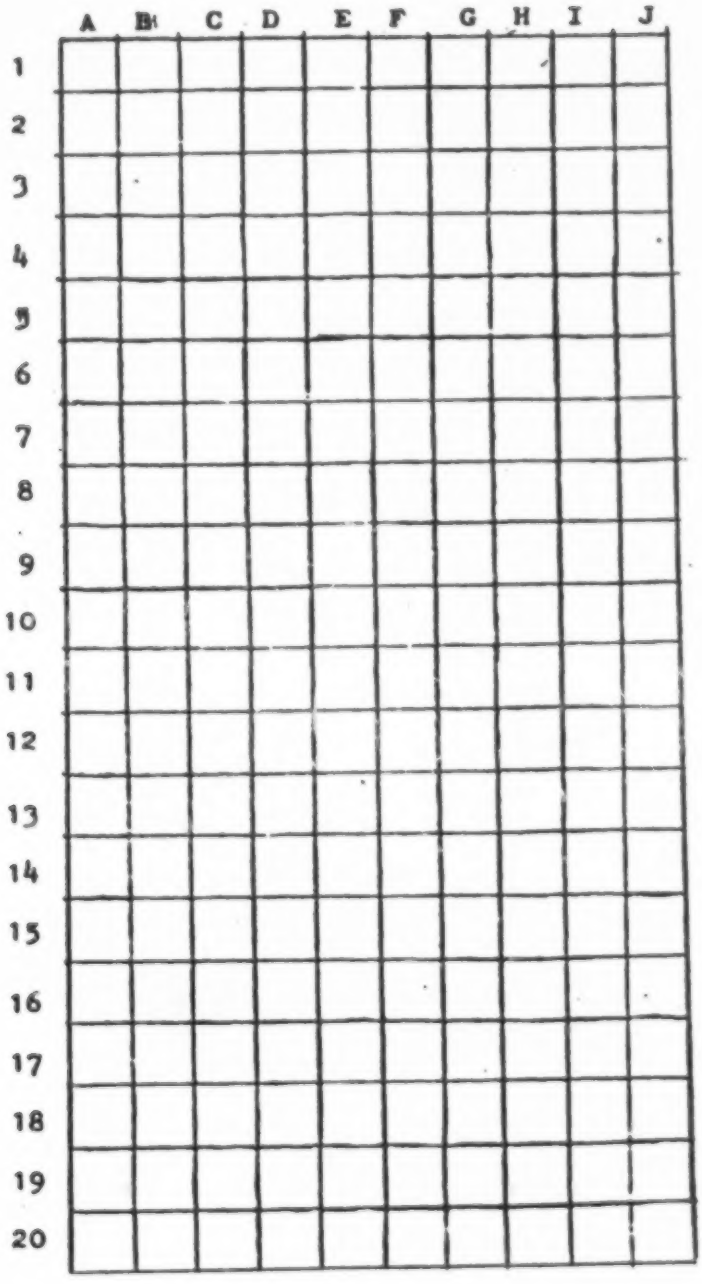
16. I shall familiar_____ you with the methods.
17. The civil servant refused to fals____ any documents.
18. Dr Thompson got the professor_____ as expected.
19. He is known as a very scholar_____ person.
20. His work had always been praise_____, therefore we could not understand why he had failed the final exam.

ANSWER SHEET
AFFIXES

Nimi: _____

Sos. turvatunnus: _____

MARK DOWN ONLY ONE ALTERNATIVE. MULTIPLE ANSWERS WILL BE COUNTED AS MISTAKES!



ADD THE SUITABLE AFFIX. MARK DOWN THE CORRECT ALTERNATIVE ON THE SEPARATE ANSWER SHEET.

a) -able b) -ance c) -ant d) -ee e) -er f) -ful g) -ic h) -ical
i) -ive j) -less

1. She is an econom(y) ____ housekeeper and feeds her family cheaply.
2. Everything was carefully planned, but even so the experiment was not success___.
3. A person who is being trained for a job is a train____.
4. In our courses attend___ is obligatory.
5. His championship is quite a respect____ achievement.

a) co- b) counter- c) equi- d) extra- e) super- f) over- g) poly-
h) sub- i) trans- j) under-

6. Things for which science currently has no explanation are _____natural.
7. The slow and steady Mr Baker is a good ____weight to the clever and unpredictable Mr Smith.
8. He got much less support than we had expected; clearly we had _____estimated him as a speaker.
9. The film that you saw was a Finnish - Swedish ____production.
10. The Americans have ____atlantic military bases in Europe.

a) anti- b) de- c) en- d) hyper- e) in- f) mis- g) out- h) post-
i) pre- j) pro-

11. He ____dated this cheque, so I cannot cash it until next week.
12. The word "battle" was _____printed as "bottle" in the newspaper
13. Cruel and unnecessary punishment ___humanizes both the punisher and the punished.
14. He was ____voted in the meeting; ten votes against him and only two in favour.
15. You will _____danger your health if you work so hard.

a) auto- b) bi- c) homo- d) inter- e) intra- f) multi- g) non-
h) semi- i) un- j) uni-

16. This is a ____weekly magazine: it is published every other week.
17. Literature other than poetry, plays, stories and novels is _____-fiction.
18. He missed the train because of a most ___fortunate accident.
19. Wages and prices depend on each other; they are ___related.
20. A person who speaks several languages is ____lingual.

END

U.S. Dept. of Education

Office of Education
Research and
Improvement (OERI)

ERIC

Date Filmed

March 29, 1991